

## ESTIMASI REGRESI WAVELET *THRESHOLDING* DENGAN METODE BOOTSTRAP

Suparti<sup>1</sup>, Achmad Mustofa<sup>2</sup> dan Agus Rusgiyono<sup>3</sup>

<sup>1,3</sup> Staf Program Studi Statistika Jurusan Matematika FMIPA UNDIP

<sup>2</sup> Alumni Jurusan Matematika FMIPA UNDIP

Jl. Prof. H. Soedarto, SH Tembalang Semarang 50275

**Abstract.** Wavelet is a function that has the certainly characteristic for example, it oscillate about zero point asscillating, localized in the time and frequency domain and construct the orthogonal bases in  $L^2(R)$  space. On of the wavelet application is to estimate non parametric regression function. There are two kinds of wavelet estimator, i.e., linear and non linear wavelet estimator. The non linear wavelet estimator is called a thresholding wavelet rstimator. The application of the bootstrap method in the thresholding wavelet function estimation is resample the wavelet coefficient of residual. The best of the thresholding wavelet estimator with bootstrap method has minimal of mean square error (MSE). The minimal MSE depend from the number of replication.

**Keywords:** non parametric regression, thresholding wavelet estimator, bootstrap method.

### 1. PENDAHULUAN

Model regresi standar dari sejumlah  $n$  data pengamatan independen

$\{(X_i, Y_i)\}_{i=1}^n$  adalah:

$$Y_i = f(X_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (1.1)$$

dengan  $X_i$  variabel prediktor,  $Y_i$  variabel respon dan  $f$  fungsi regresi yang tidak diketahui. Sementara  $\varepsilon_i$  variabel random independen dengan mean 0 dan varian  $\sigma^2$ . Ada dua pendekatan dalam mengestimasi fungsi  $f$  yaitu pendekatan parametrik dan pendekatan non-parametrik [1]. Pendekatan parametrik dilakukan jika asumsi bentuk fungsi  $f$  diketahui tergantung dari suatu parameter misalnya linear, eksponensial, dan lain-lain, sedangkan pendekatan non-parametrik dilakukan jika asumsi bentuk  $f$  tidak diketahui. Teknik regresi non-parametrik seperti metode kernel dan deret Fourier mengasumsikan bahwa fungsi  $f$  termuat dalam kelas fungsi mulus, artinya mempunyai turunan yang kontinu. Jika fungsinya tidak mulus maka kedua metode tersebut tidak dapat digunakan dengan baik. Sedangkan dengan metode wavelet asumsi kemulusan fungsi dapat diperlemah, karena wavelet mampu mengestimasi baik fungsi mulus maupun tidak mulus.

Estimator wavelet sendiri dibedakan menjadi dua macam, yaitu estimator wavelet linear dan estimator wavelet non-linear. Estimator wavelet nonlinear dinamakan juga *estimator wavelet thresholding* atau *estimator wavelet shrinkage*. Salah satu ukuran kebaikan dari estimator tersebut adalah nilai Mean Square Error (MSE) atau nilai Error Kuadrat Rata-rata Terintegrasi / Integrated Mean Square Error (IMSE). Estimator wavelet linear mempunyai penurunan IMSE lebih cepat menuju nol dari pada estimator deret Fourier tetapi sama cepatnya dengan estimator kernel [2]. Sedangkan estimator nonlinear mempunyai laju konvergensi IMSE yang lebih cepat menuju nol dari estimator linear [3].

Bootstrap merupakan prosedur untuk mendapatkan estimasi parameter, dengan resampling data dengan pengembalian. Efron dan Tibshirani [4] menggunakan metode bootstrap dalam menentukan estimasi fungsi regresi linear dengan resampling residual. Sedangkan dalam makalah ini metode bootstrap dalam estimasi fungsi *regresi wavelet thresholding* dengan melakukan resampling koefisien wavelet dari residual seperti yang dilakukan oleh Bruce dan Gao [5].

### 1.1. Estimator deret Fourier

Diasumsikan bahwa  $f \in L^2(R)$  dengan  $L^2(R) = \left\{ f : \int_{-\infty}^{\infty} f(x)^2 dx < \infty \right\}$ , maka  $L^2(R)$  merupakan ruang Hilbert [6]. Sebuah hasil kali dalam pada ruang  $L^2(R)$  adalah fungsi yang mengasosiasikan bilangan riil  $\langle f, g \rangle$ , dengan masing-masing pasangan fungsi  $f(x)$  dan  $g(x)$  pada  $L^2(R)$ . Hasil kali dalam  $L^2(R)$  dari dua fungsi dan norma sebuah fungsi didefinisikan

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)g(x)dx, \\ \|f\| = \sqrt{\langle f, f \rangle} = \sqrt{\int_{-\infty}^{\infty} (f(x))^2 dx}.$$

Andaikan  $\{\phi_j\}_{j=1,2,\dots}$  sistem ortonormal lengkap (CONS) dari  $L^2(R)$ , maka sembarang  $f \in L^2(R)$  dapat dinyatakan sebagai  $f = \sum_{j=1}^{\infty} \alpha_j \phi_j$  dengan

$$\alpha_j = \langle f, \phi_j \rangle,$$

dan memenuhi identitas Parseval

$$\|f\|^2 = \sum_{j=1}^{\infty} \alpha_j^2. \quad \text{Karena} \quad \int_{-\infty}^{\infty} f(x)^2 dx < \infty$$

maka  $\sum_{j=1}^n \alpha_j^2 < \infty$  sehingga  $\alpha_j \rightarrow 0$ , untuk  $j \rightarrow \infty$ . Oleh karena itu,  $f$  dapat didekati

oleh  $f = \sum_{j=1}^J \alpha_j \phi_j$  untuk bilangan bulat  $J$

cukup besar. Khususnya jika  $f \in L^2[0, 2\pi]$ , maka  $f$  dapat didekati dengan deret Fourier,

$$f_J(x) = \frac{1}{2} a_0 + \sum_{j=1}^J (a_j \cos(jx) + b_j \sin(jx)), \quad (1.2)$$

dengan koefisien Fourier

$$a_j = \frac{1}{\pi} \langle f, \cos(j \cdot) \rangle = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(jx) dx$$

dengan  $j=0, 1, \dots, J$  dan

$$b_j = \frac{1}{\pi} \langle f, \sin(j \cdot) \rangle = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(jx) dx$$

dengan  $j=1, 2, \dots, J$ .

Jika  $\{(X_i, Y_i)\}_{i=1}^n$  merupakan data observasi independen mempunyai model

(1.1) dengan  $X_i = \frac{2\pi i}{n}$  dan  $X_i \in [0, 2\pi]$ , maka estimator regresi  $f$  adalah

$$\hat{f}_J(x) = \frac{1}{2} \hat{a}_0 + \sum_{j=1}^J (\hat{a}_j \cos(jx) + \hat{b}_j \sin(jx)), \quad (1.3)$$

dengan  $\hat{a}_j = \frac{2}{n} \sum_{i=1}^n Y_i \cos(jX_i), j = 0, 1, \dots, J$

dan  $\hat{b}_j = \frac{2}{n} \sum_{i=1}^n Y_i \sin(jX_i), j = 1, 2, 3, \dots, J$ .

### 1.2. Fungsi Wavelet

Fungsi wavelet adalah suatu fungsi dengan sifat-sifat tertentu diantaranya yang berosilasi di sekitar nol (seperti fungsi sinus dan cosinus), terlokalisasi dalam domain waktu dan frekuensi serta membentuk basis ortogonal dalam  $L^2(R)$  [7]. Fungsi wavelet dibedakan atas dua jenis, yaitu wavelet ayah ( $\phi$ ) dan wavelet ibu ( $\psi$ ) yang mempunyai sifat:

$$\int_{-\infty}^{\infty} \phi(x) dx = 1 \quad \text{dan} \quad \int_{-\infty}^{\infty} \psi(x) dx = 0.$$

Dengan dilatasi diadik dan translasi integer, wavelet ayah dan wavelet ibu melahirkan keluarga wavelet yaitu

$$\phi_{j,k}(x) = (p2^j)^{1/2} \phi(p2^j x - k) \quad \text{dan}$$

$$\psi_{j,k}(x) = (p2^j)^{1/2} \psi(p2^j x - k),$$

untuk suatu skalar  $p > 0$ , dan tanpa mengurangi keumuman dapat diambil  $p=1$ , sehingga

$$\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k) \quad \text{dan}$$

$$\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k).$$

Fungsi  $\phi_{j,k}(x)$  dan  $\psi_{j,k}(x)$  mempunyai sifat

$$\int_{-\infty}^{\infty} \phi_{j,k}(x) \phi_{j',k'}(x) dx = \delta_{k,k'},$$

$$\int_{-\infty}^{\infty} \psi_{j,k}(x) \phi_{j',k'}(x) dx = 0,$$

$$\int_{-\infty}^{\infty} \psi_{j,k}(x) \psi_{j',k'}(x) dx = \delta_{j,j'} \delta_{k,k'},$$

dengan  $\delta_{i,j} = \begin{cases} 1 & \text{jika } i = j \\ 0 & \text{jika } i \neq j. \end{cases}$

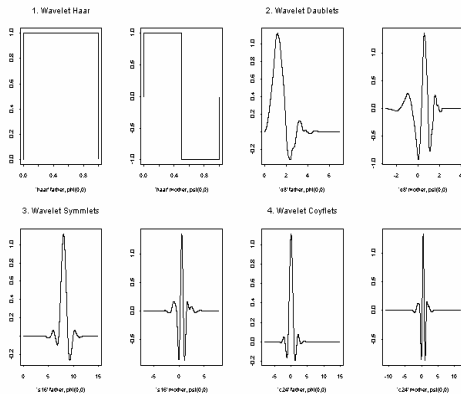
Contoh wavelet paling sederhana adalah wavelet Haar yang mempunyai rumus

$$\psi(x) = \begin{cases} 1, & 0 \leq x \leq 1/2 \\ -1, & 1/2 \leq x < 1 \\ 0, & x \text{ yang lain} \end{cases}$$

dan

$$\phi(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & x \text{ yang lain} \end{cases}$$

Gambar 1 adalah beberapa contoh wavelet yang meliputi wavelet Haar, wavelet Daubechies (Daublet), symmetris (Symmlet), dan Coifman (Coiflet) [8].



Gambar 1. Beberapa contoh wavelet

## 2. ANALISIS MULTIRESOLUSI

Analisis multiresolusi  $L^2(\mathbb{R})$  adalah ruang bagian tertutup  $\{V_j, j \in \mathbb{Z}\}$  yang memenuhi

- $\dots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \dots$
- $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}, \bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R})$
- $f \in V_j \Leftrightarrow f(2 \cdot) \in V_{j+1}$
- $f \in V_0 \Rightarrow f(-k) \in V_0, \forall k \in \mathbb{Z}$
- Terdapat sebuah fungsi  $\phi \in V_0$  sehingga  $\phi_{0,k} = \phi(\cdot - k), k \in \mathbb{Z}$  membentuk basis ortonormal untuk  $V_0$  dimana untuk semua  $j, k \in \mathbb{Z}$ ,  

$$\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k).$$

Jika  $\{V_j, j \in \mathbb{Z}\}$  analisis multiresolusi dari  $L^2(\mathbb{R})$ , maka ada basis ortonormal  $\{\psi_{j,k}; j, k \in \mathbb{Z}\}$  untuk  $L^2(\mathbb{R})$ :

$$\psi_{j,k} = 2^{j/2} \psi(2^j x - k),$$

sehingga untuk sembarang  $f$  pada  $L^2(\mathbb{R})$ ,

$$P^j f = P^{j-1} f + \sum_{k \in \mathbb{Z}} \langle f, \psi_{j-1,k} \rangle \psi_{j-1,k},$$

yaitu  $\psi(x)$  yang diturunkan dari

$$\psi(x) = \sum_{k \in \mathbb{Z}} (-1)^k c_{(-k+1)} \phi_{1,k}(x).$$

**Akibat.** Bila  $\phi$  adalah fungsi skalar yang membangun analisis multiresolusi dan

$$\psi(x) = \sum_{k \in \mathbb{Z}} (-1)^k c_{(-k+1)} \phi_{1,k}(x),$$

maka dekomposisi sembarang  $f \in L^2(\mathbb{R})$  ke dalam wavelet ortonormal adalah

$$f(x) = \sum_{k \in \mathbb{Z}} c_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x), \quad (2.1)$$

dengan

$$c_{j_0,k} = \langle f, \phi_{j_0,k} \rangle \text{ dan } d_{j,k} = \langle f, \psi_{j,k} \rangle.$$

## 3. ESTIMATOR WAVELET LINEAR

Misalkan terdapat sekumpulan data independen  $\{(X_i, Y_i)\}_{i=1}^n$  yang mempunyai model (1.1) dan  $n = 2^m$  dengan  $m$  bilangan bulat positif. Jika  $X_i$  rancangan titik reguler pada interval  $[0, 1]$  dengan  $X_i = i/n$ , maka proyeksi  $f$  pada ruang  $V_J$  dapat ditulis menjadi

$$(P^J f) = \sum_{k \in \mathbb{Z}} c_{J,k} \phi_{J,k}(x) \text{ atau}$$

$$f_J(x) = \sum_{k \in \mathbb{Z}} c_{J,k} \phi_{J,k}(x),$$

dengan  $c_{J,k} = \langle f, \phi_{J,k} \rangle = \int_0^1 f(x) \phi_{J,k}(x) dx$ .

Untuk  $J \rightarrow \infty$  maka  $f_J(x) \rightarrow f(x)$ . Berdasarkan dekomposisi fungsi ke dalam wavelet ortonormal (2.1) untuk sembarang fungsi  $f \in L^2(\mathbb{R})$  diperoleh

$$f_J(x) = \sum_{k \in \mathbb{Z}} c_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x),$$

dengan

$$c_{j_0,k} = \langle f, \phi_{j_0,k} \rangle = \int_0^1 f(x) \phi_{j_0,k}(x) dx \text{ dan}$$

$$d_{j,k} = \langle f, \psi_{j,k} \rangle = \int_0^1 f(x) \psi_{j,k}(x) dx.$$

Karena fungsi regresi  $f$  tidak diketahui maka estimator  $\hat{f}$  pada ruang  $V_J$  dapat ditulis sebagai

$$\hat{f}_J(x) = \sum_{k \in Z} \hat{c}_{J,k} \phi_{J,k}(x)$$

dengan  $\hat{c}_{J,k} = \frac{1}{n} \sum_{i=1}^n Y_i \phi_{J,k}(X_i)$ , atau

$$\hat{f}_J(x) = \sum_{k \in Z} \hat{c}_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k \in Z} \hat{d}_{j,k} \psi_{j,k}(x) \quad (3.1)$$

dengan

$$\hat{c}_{j_0,k} = \frac{1}{n} \sum_{i=1}^n Y_i \phi_{j_0,k}(X_i),$$

$$\hat{d}_{j,k} = \frac{1}{n} \sum_{i=1}^n Y_i \psi_{j,k}(X_i).$$

Estimator wavelet (5) dinamakan estimator wavelet linear.

#### 4. ESTIMATOR WAVELET THRESHOLDING

Jika diberikan data  $\{(X_i, Y_i)\}_{i=1}^n$  dengan model (1),  $n = 2^m$  dan  $X_i = i/n$ , maka  $Y_i \sim N(g(i/n), \sigma^2)$ . Mean dan varian dari  $\hat{d}_{j,k}$  adalah  $E[\hat{d}_{j,k}] = d_{j,k}$  dan  $\text{Var}(\hat{d}_{j,k}) = \frac{\sigma^2}{n}$  sehingga  $\hat{d}_{j,k} \sim N(d_{j,k}, \frac{\sigma^2}{n})$ .

Jadi koefisien wavelet empiris  $\hat{d}_{j,k}$  memuat sejumlah noise dan hanya relatif sedikit yang memuat sinyal signifikan. Oleh karena itu, dapat direkonstruksi estimator wavelet dengan menggunakan sejumlah koefisien terbesar [7,9]. Yakni hanya koefisien yang lebih besar dari suatu nilai tertentu yang diambil, sedangkan koefisien selebihnya diabaikan, karena dianggap 0. Nilai tertentu tersebut dinamakan nilai *threshold* (nilai ambang) dan estimatornya menghasilkan

$$\hat{f}_\lambda(x) = \sum_k \hat{c}_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k=0}^{2^j-1} \partial_\lambda(\hat{d}_{j,k}) \psi_{j,k}(x) \quad (4.1)$$

dengan  $\partial_\lambda$  menyatakan fungsi *thresholding* atau fungsi ambang dengan nilai ambang atau *threshold*  $\lambda$ . Estimator (6) dinamakan estimator wavelet non linear, *estimator*

*wavelet shrinkage*, atau *estimator wavelet thresholding*.

Karena *thresholding* dirancang untuk membedakan antara koefisien wavelet empiris yang masuk dan yang keluar dari rekonstruksi wavelet, sedangkan untuk membuat keputusan ada 2 faktor yang mempengaruhi ketepatan estimator, yaitu ukuran sampel  $n$  dan tingkat noise  $\sigma^2$ , maka setiap koefisien merupakan calon kuat masuk di dalam rekonstruksi wavelet jika ukuran sampel besar atau tingkat noise kecil. Karena  $\sqrt{n} \hat{d}_{j,k} / \sigma$  berdistribusi normal dengan varian 1 untuk seluruh  $n$  dan  $\sigma$ , maka estimator *thresholding* dari  $d_{j,k}$  adalah

$$\tilde{d}_{j,k} = \frac{\sigma}{\sqrt{n}} \partial_\lambda \left( \frac{\sqrt{n} \hat{d}_{j,k}}{\sigma} \right),$$

sehingga *estimator wavelet thresholding*-nya adalah

$$\hat{f}_\lambda(x) = \sum_k \hat{c}_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k=0}^{2^j-1} \frac{\sigma}{\sqrt{n}} \partial_\lambda \left( \frac{\sqrt{n} \hat{d}_{j,k}}{\sigma} \right) \psi_{j,k}(x), \quad (4.2)$$

dengan

$$\begin{aligned} \hat{c}_{j_0,k} &: \text{penduga dari } c_{j_0,k} \\ \hat{d}_{j,k} &: \text{penduga dari } d_{j,k} \\ \lambda &: \text{parameter threshold} \\ \partial_\lambda &: \text{fungsi thresholding} \end{aligned}$$

#### Langkah-langkah Thresholding

Langkah-langkah *thresholding* adalah sebagai berikut.

##### 1. Pemilihan Fungsi Thresholding

Ada dua jenis fungsi *thresholding*  $\partial_\lambda$ , yaitu:

###### a. Hard Thresholding,

$$\partial_\lambda^H(x) = \begin{cases} x, & |x| > \lambda \\ 0, & \text{x yang lain} \end{cases}$$

###### b. Soft Thresholding,

$$\partial_\lambda^S(x) = \begin{cases} x - \lambda, & x > \lambda \\ 0, & x \leq \lambda \\ x + \lambda, & x < -\lambda \end{cases}$$

dengan  $\lambda$  merupakan parameter *threshold*.

Fungsi Hard *thresholding* lebih dikenal karena terdapat diskontinuitas dalam fungsi *thresholding* sehingga nilai  $x$  yang berada di atas *threshold*  $\lambda$  tidak disentuh. Sebaliknya, fungsi soft *thresholding* kontinu yaitu sejak nilai  $x$  berada di atas *threshold*  $\lambda$ . Motivasi penggunaan soft *thresholding* berasal dari prinsip bahwa noise mempengaruhi seluruh koefisien wavelet. Juga kekontinuitas dari fungsi soft shrinkage membuat kondisi yang lebih baik untuk alasan statistik.

## 2. Estimasi $\sigma$

Dalam merekonstruksi fungsi wavelet biasanya nilai  $\sigma$  tidak diketahui. Oleh karena itu,  $\sigma$  harus diestimasi dari data. Ogden [7] memberikan estimasi  $\sigma$  berdasarkan koefisien wavelet empiris pada level resolusi tertinggi dengan fungsi Median Deviasi Absolut (MAD), yaitu:

$$\hat{\sigma} = \frac{\text{median}(\hat{d}_{J-1,k} - \text{median}(\hat{d}_{J-1,k}))}{0,6745}.$$

## 3. Pemilihan Parameter *Threshold*

Pada estimasi fungsi dengan metode wavelet *thresholding*, tingkat kemulusan estimator ditentukan oleh level resolusi  $J$ , fungsi *thresholding*  $\partial_\lambda$  dan parameter *threshold*  $\lambda$ . Namun pemilihan  $J$  dan  $\partial_\lambda$  tidak sedominan  $\lambda$ . Nilai  $\lambda$  yang terlalu kecil memberikan estimasi fungsi yang sangat tidak mulus (*under smooth*) sedangkan nilai  $\lambda$  yang terlalu besar memberikan estimasi yang sangat mulus (*over smooth*). Oleh karena itu perlu dipilih parameter *threshold* yang optimal untuk mendapatkan fungsi yang optimal. Untuk memilih nilai *threshold* optimal, ada dua kategori pemilihan yaitu memilih satu harga *threshold* untuk seluruh level resolusi (pemilihan secara global) dan pemilihan *threshold* yang tergantung pada level resolusi.

Untuk pemilihan global *threshold*, Ogden [7] memberikan 2 pemilihan *threshold* yang hanya bergantung pada banyaknya data pengamatan  $n$  yaitu

*threshold* universal ( $\lambda_j = \sqrt{2 \log n}$ )

dan *threshold* minimax yang telah ditabelkan oleh Donoho dan Johnstone [10]. Nilai-nilai *threshold* minimax selalu lebih kecil dibandingkan dengan nilai *threshold* universal untuk ukuran sampel yang sama.

Pemilihan nilai *threshold* berdasarkan level resolusi memberikan kemungkinan adanya perbedaan nilai *threshold*  $\lambda_j$  yang dipilih untuk tiap level resolusi  $j$ . Ada beberapa cara pemilihan *threshold* yang tergantung pada level resolusi, diantaranya *threshold* Adapt dan *threshold* Top.

*Threshold* adapt didasarkan pada prinsip untuk meminimalkan *Stein Unbiased Risk Estimator* (SURE) pada suatu level resolusi. *Threshold* adapt untuk himpunan koefisien detail  $d_j$  yang beranggotakan  $K$  koefisien didefinisikan sebagai

$$\lambda_j = \arg \min_{t \geq 0} \text{SURE}(d_j, t),$$

dengan

$$\text{SURE}(d_j, t) = K - 2 \sum_{k=1}^K 1_{\{|d_{j,k}| \leq t \sigma_j\}} + \sum_{k=1}^K \min \left\{ (d_{j,k} / \sigma_j)^2, t^2 \right\}$$

Sedangkan nilai *threshold* Top ditentukan berdasarkan besar prosentase koefisien yang akan digunakan dari keseluruhan koefisien wavelet dalam merekonstruksi fungsi.

## 4.1 Estimator Regresi Wavelet *Thresholding* Terbaik dengan Metode Bootstrap

Suatu kebaikan estimator dapat dilihat dari besarnya tingkat kesalahan. Semakin kecil tingkat kesalahannya semakin baik estimatornya. Salah satu ukuran kebaikan suatu estimator adalah MSE. Suatu estimator  $\hat{f}(x)$  mempunyai  $\text{MSE}(\hat{f}(x)) = \text{var}(\hat{f}(x)) + \text{bias}^2(\hat{f}(x))$  dengan  $\text{bias}(\hat{f}(x)) = E(\hat{f}(x) - f(x))$ . Karena  $f(x)$  tidak diketahui maka nilai MSE tidak dapat diketahui juga, sehingga perlu dilakukan estimasi MSE. Abramovich dan

Benjamini [11] memberikan estimasi dari MSE sebagai

$$MSE = n^{-1} \|y - \hat{y}\|^2 = n^{-1} \sum_{i=1}^n (y - \hat{y})^2.$$

Bootstrap merupakan suatu metode resampling dengan pengembalian. Bootstrap dalam estimasi regresi dapat dilakukan melalui resampling pada data, residual atau yang lain. Dalam makalah ini bootstrap dilakukan dengan meresampling koefisien wavelet dari residual. Untuk mendapatkan estimator regresi wavelet *thresholding* terbaik dengan metode bootstrap, jika dibangkitkan sampel bootstrap sebanyak  $M$  kali, maka akan didapatkan Mean Square Error (MSE) sebanyak  $M$ . Dari sebanyak  $M$  MSE ini dipilih MSE minimal. Estimator yang meminimalkan MSE ini merupakan estimator terbaik dari  $M$  resampling bootstrap. Secara garis besar langkah-langkah untuk menentukan estimator regresi wavelet *thresholding* dengan metode bootstrap sebagai berikut.

1. Melakukan estimasi fungsi regresi wavelet *thresholding*  $\hat{Y}$  dengan prosedur wavelet *shrinkage* standar.
2. Menghitung residual dari estimasi wavelet *thresholding*.
3. Menghitung koefisien dari residual
4. Membentuk koefisien wavelet baru dengan melakukan resampling bootstrap pada koefisien wavelet residual secara sendiri-sendiri atau bersama-sama.
5. Menyusun data baru  $Y^{*(i)}$  berdasarkan bootstrap koefisien wavelet residual.
6. Merekonstruksi estimasi regresi wavelet *thresholding*  $\hat{Y}^{(i)}$  berdasarkan sampel bootstrap  $Y^{*(i)}$ .
7. Menghitung Mean Square Error (MSE).
8. Mengulangi langkah 4 sampai dengan 7 sebanyak  $M$  kali sehingga diperoleh  $M$  estimasi fungsi regresi bootstrap  $\hat{Y}^{(1)}, \hat{Y}^{(2)}, \hat{Y}^{(3)}, \dots, \hat{Y}^{(M)}$ .
9. Memilih  $\hat{Y}^{(i)}$  dengan MSE terkecil.
10. Gambar *estimator wavelet thresholding* terbaik.

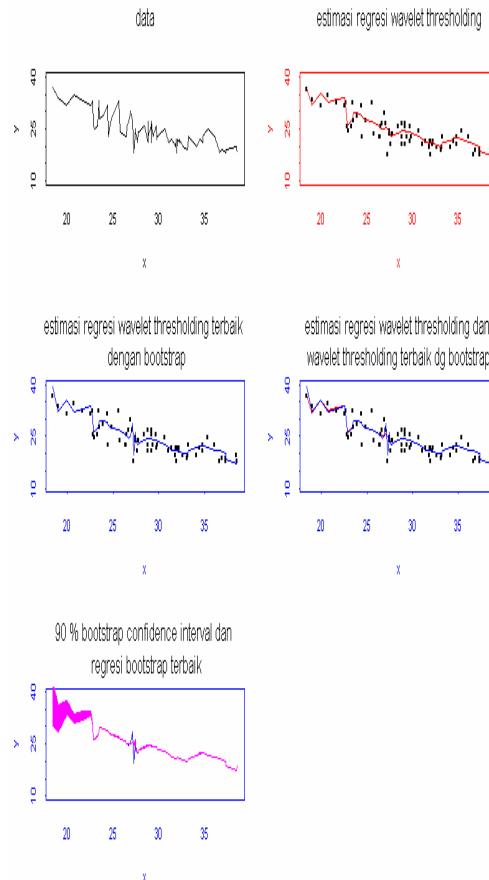
## 4.2 Studi kasus

Untuk menerapkan metode bootstrap dalam estimasi regresi wavelet digunakan data berat dan jumlah gas permil model automobil 1990 [12]. Dalam hal ini variabel-variabelnya sebagai berikut.

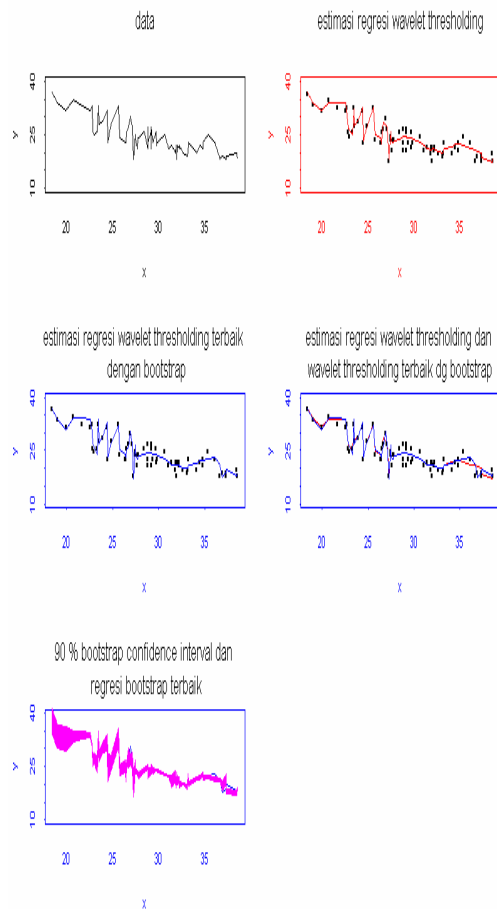
- Sebagai variabel respon,  $Y$  menyatakan jumlah gas permil
- Sebagai variabel prediktor,  $X$  (dalam kuintal) menyatakan berat automobil

Dari data tersebut dicari hubungan antara  $Y$  dan  $X$  yaitu mencari estimasi kurva regresi wavelet *thresholding* dengan metode bootstrap terbaik. Berikut adalah estimasi dengan *thresholding* universal dan *minimax* menggunakan program S-PLUS

+ Wavelets



Gambar 2. Estimasi Wavelet *Thresholding* Menggunakan *Threshold* Universal dengan Prosedur Bootstrap



Gambar 3. Estimasi Wavelet *Thresholding* Menggunakan *Threshold* Minimax dengan Prosedur Bootstrap

Keterangan Gambar 2 dan Gambar 3.

- ~~~~~ / ..... : data
- (red) : estimasi regresi wavelet *thresholding*
- (blue) : estimasi regresi wavelet *thresholding* terbaik dengan bootstrap
- (magenta) : pendekatan 90% interval konfidensi dengan bootstrap

Dari Gambar 2, estimasi wavelet *thrsholding* dengan bootstrap terbaik dan tanpa bootstrap menghasilkan kurva yang hampir sama. Dari 120 pengulangan, 90 persennya sebagian besar masih berada pada garis-garis tebal, dan titik-titik yang berada pada garis yang saling berdekatan

merupakan nilai-nilai yang dekat dengan estimasi, tapi ada titik-titik yang diluar estimasinya. Dari data sebanyak  $n = 66$  dan resampling sebanyak 120 kali didapatkan MSE wavelet *thresholding* tanpa bootstrap sebesar 5.43752567186436 dan MSE minimal setelah pembootstrapan sebesar 4.88123820765525 yang terletak pada sampel ke -113

Dari Gambar 3, estimasi wavelet *thresholding* dengan *threshold* minimax dengan bootstrap terbaik dan tanpa bootstrap dihasilkan kurva yang hampir sama. Dari 120 pengulangan 90 persennya sebagian besar masih berada pada garis-garis tebal, dan titik-titik yang berada pada garis yang saling berdekatan merupakan nilai-nilai yang dekat dengan estimasi, tapi ada titik-titik yang diluar estimasinya. Dari data sebanyak  $n = 66$  dan resampling sebanyak 120 kali didapatkan MSE wavelet *thresholding* tanpa bootstrap sebesar 2.36521414391298 dan MSE minimal setelah pembootstrapan sebesar 2.12811836696285 yang dicapai pada sampel ke -84 .

Dari pengestimasian kurva dengan metode bootstrap menggunakan *threshold* universal dan minimax, MSE minimal dengan menggunakan *threshold* universal minimax lebih kecil dari MSE minimal dengan *threshold* universal.

## 5. KESIMPULAN

Pada estimasi wavelet *thresholding* dengan bootstrap, pemilihan parameter *threshold* dan banyaknya replikasi bootstrap berpengaruh pada hasil MSE. Penggunaan parameter *threshold* minimax menghasilkan MSE yang lebih kecil dari pada *threshold* universal.

## 6. DAFTAR PUSTAKA

- [1] Abramovich, F. and Benjamini, Y. (1995), *Thresholding of Wavelet Coefficients as Multiple Hypothesis Testing Procedure In Wavelets and Statistics*, Springer-Verlag, New York, 5-14.

- [2] Bruce, A. and Gao, H Y. (1996), *Applied Wavelet Analisis with S-PLUS*, Springer-Verlag. New York.
  - [3] Daubechies, I. (1992), *Ten Lectures on Wavelets*, Capital City Press, Philadelphia.
  - [4] Donoho, D.L and Johnstone, I.M. (1994), *Ideal Spatial Adaptation by Wavelet Shrinkage*, Biometrika, **81** (3), 425-455.
  - [5] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., and Picard, D. (1996), *Density Estimation by Wavelet Thersholding*. The Annals of Statistics, **24**(2), 508-539.
  - [6] Efron, B., Tibshirani. (1993), *An Introduction to the Bootstrap*, Chapman and Hall. New York.
  - [7] Hall, P and Patil. P. (1995), *On Wavelet Methods for Estimating Smooth Function*, Bernoulli 1(1/2). 041-058.
  - [8] Hardle, W. (1993), *Applied Non-parametric Regression*. Cambridge University Press. New York.
  - [9] [Http://www.en.wikipedia.org/wiki/Illustration\\_of\\_density\\_nonparametric](http://www.en.wikipedia.org/wiki/Illustration_of_density_nonparametric)
  - [10] Ogden, R.T. (1997), *Essential Wavelet for Statistical Application and Data Analysis*, Birkhauser, Boston.
  - [11] Suparti dan Subanar, H. (2000), *Estimasi Regresi dengan Metode Wavelet Shrinkage*. Jurnal Sains & Matematika, **8**(3), 105-113.
  - [12] Vetterli, M. and Kovacevic, J. (1995), *Wavelets And Subband Co-ding*. Prentice Hall PTR, New Jersey.
-